A Comparison Of Symptom Validity Tests For Academic Neuropsychological Evaluations: Medical Symptom Validity Test Versus The Word Reading Test

Troy J. Stettler
Pacific University

Recommended Citation
A Comparison Of Symptom Validity Tests For Academic Neuropsychological Evaluations: Medical Symptom Validity Test Versus The Word Reading Test

Abstract
Symptom Validity Tests (SVTs) are designed to detect suboptimal effort during neuropsychological testing. A majority of SVTs, including the Medical Symptom Validity Test (MSVT), employ memory test formats and are designed for use among patient groups where memory complaints are prominent. In contrast, the Word Reading Test (WRT) is designed to assess effort specific to reading, a common complaint among patients undergoing neuropsychological evaluation for academic purposes. The WRT and the MSVT were administered as part of cognitive evaluations at a university doctoral clinical psychology training and research clinic. Of the 30 cases analyzed, six (20%) failed either the WRT, the MSVT, or both. The present study supports a "general-global" hypothesis of effort. Clinical implications of the findings are discussed.

Degree Type
Thesis

Degree Name
Master of Science in Psychology

Committee Chair
Michael Daniel, Ph. D.

Subject Categories
Psychiatry and Psychology

This thesis is available at CommonKnowledge: https://commons.pacificu.edu/spp/51
A COMPARISON OF SYMPTOM VALIDITY TESTS FOR ACADEMIC
NEUROPSYCHOLOGICAL EVALUATIONS:
MEDICAL SYMPTOM VALIDITY TEST VERSUS THE WORD READING TEST

A THESIS
SUBMITTED TO THE FACULTY
OF
SCHOOL OF PROFESSIONAL PSYCHOLOGY
PACIFIC UNIVERSITY
HILLSBORO, OREGON
BY
TROY J STETTLER
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
OF
MASTER OF SCIENCE IN CLINICAL PSYCHOLOGY
April 17, 2009

APPROVED: ___________________________ Michael Daniel, Ph.D.
ABSTRACT

Symptom Validity Tests (SVTs) are designed to detect suboptimal effort during neuropsychological testing. A majority of SVTs, including the Medical Symptom Validity Test (MSVT), employ memory test formats and are designed for use among patient groups where memory complaints are prominent. In contrast, the Word Reading Test (WRT) is designed to assess effort specific to reading, a common complaint among patients undergoing neuropsychological evaluation for academic purposes. The WRT and the MSVT were administered as part of cognitive evaluations at a university doctoral clinical psychology training and research clinic. Of the 30 cases analyzed, six (20%) failed either the WRT, the MSVT, or both. The present study supports a “general-global” hypothesis of effort. Clinical implications of the findings are discussed.
TABLE OF CONTENTS

Page
ABSTRACT ................................................................................................................................. ii
LIST OF TABLES .................................................................................................................... iv
INTRODUCTION ..................................................................................................................... 1
REVIEW OF THE LITERATURE .............................................................................................. 7
EFFORT TESTING AND INCENTIVES ..................................................................................... 7
  Malingering ............................................................................................................................. 9
  Base rates ............................................................................................................................. 10
EFFORT INDICATORS AND SYMPTOM VALIDITY MEASURES ........................................ 12
  Patient presentation and history ......................................................................................... 12
  Indices within neuropsychological and psychological tests. ............................................. 13
  Development and clinical use of symptom validity measures .......................................... 14
  Independent Symptom Validity Tests. .................................................................................. 23
Hypotheses of the Study ......................................................................................................... 27
METHOD ................................................................................................................................. 24
  Procedure ............................................................................................................................. 25
RESULTS .................................................................................................................................. 27
DISCUSSION ............................................................................................................................ 39
REFERENCES .......................................................................................................................... 42
List of Tables

Table 1. Age, Education and WAIS-III Indices of the Sample ............................................25

Table 2. Group Means and Standard Deviations for Percent Correct on the MSVT Subtests. .................................................................................................................................28

Table 3. Chi-Square Results for those failing the Medical Symptoms Validity Test........29

Table 4. Analysis of Variance for Effort Test Failure the Medical Symptom .................

Table 5. Chi-Square Results for Effort Test Failure on the Word Reading Test ............32

Table 6. Analysis of Variance for Effort Test Failure on the Word Reading Test ..........33

Table 7. Effort Failure Overlap with Cutoff Scores for the WRT Mean Error ..............34

Table 8. ANOVA Results for Errors and Reaction Time on the Word Reading Test......35

Table 9. Chi-Square Results for Effort Test Failure on either the MSVT or the WRT ....36

Table 10. Analysis of Variance for Effort Test Failure on Either the MSVT or the WRT ..............................................................................................................................37

Table 11. Correlations between MSVT and WRT Scores .............................................38
INTRODUCTION

Considerable recent neuropsychological research has focused on effort testing. Effort testing is designed to assess whether or not an examinee gives sufficient effort during a neuropsychological evaluation to obtain valid test results. The importance of effort testing to neuropsychologists and their related medical and forensic colleagues was emphasized by Richman et al. (2006). The author’s state:

Undetected symptom exaggeration is not just of economic interest to insurance companies, although that is a consideration with major financial implications. It could also be a major contaminating variable in clinical studies of the relative effectiveness of alternative forms of treatment to the extent that self-reported function or symptom reporting is assessed. Exaggerated symptoms could lead to unnecessary treatments and possibly, in some cases, to severe adverse side effects, especially when there are no objective medical findings and the diagnosis relies on self reported symptoms. (p. 310)

Lezak, Howieson, and Loring (2004) noted that many factors are known to affect neuropsychological test scores. It now seems effort may contribute more variance in test scores than any other variable. In fact, Constantinou, Bauer, Ashendor, Fisher, and McCaffery (2005) found that among litigating patients, 47% of the variance on the General Neuropsychological Deficit Scale is accounted for by effort test results. Rohling, Green, Allen, and Less-Haley (2000) found similar results when analyzing 657 patents seen for disability evaluations with an overall correlation between measures of cognitive performance and symptom exaggeration of .72. Their results showed a stronger relationship between effort and cognitive scores than brain injury severity and cognitive scores. To better understand the effects of effort and external incentives, Flaro, Green,
and Roberson (2007) showed failure on an effort test was twice as frequent in a mild Traumatic Brain Injury (TBI) group that was seeking compensation or involved in litigation than those in a severe TBI group. The authors concluded that differences in failure rates on effort tests cannot be explained by differences in brain injury severity yet they are explainable by differences in external incentives.

While these results are notable, it is important to keep in mind that the research mentioned above was conducted within forensic populations, specifically with patients seeking financial compensation (for similar results, see Green, 2003; Green, 2007; Green, Rohling, Lees-Haley, & Allen, 2000; Hartman, 2002). Many, if not most, neuropsychological assessments are conducted for reasons other than litigation. However, effort among other populations, such as university settings, is still a substantial concern (Osmon, Plambeck, Klein, & Mano, 2006).

For example, researchers and clinicians have described incentives for poor effort among patients undergoing Learning Disability (LD) and Attention Deficit Hyperactivity Disorder (ADHD) evaluations. Harrison (2006) noted students with LD and ADHD may receive assistive devices, tutoring services, individual room campus housing, dismissal of student loans, and government sponsored bursary program (Canada only) that will allow them to purchase computers. Sullivan, May, & Galbally (2007) identified other accommodations such as alternative courses and psychostimulant medication, which have a history of misuse, abuse, and distribution among college students (for a review of psychostimulant misuses among college students see Barrett, Darredeau, Bordy, & Pihl, 2005; Upadhaya, et al., 2005). Some researchers have suggested a primary incentive for
feigning effort is to gain extra time on college entrance examinations. Mullins (2003) argues:

In an era of competitive admissions and over diagnosis of attention disorders, educators worry that high school students (and their parents) will exaggerate or falsify claims of attention-deficit disorder to gain a competitive edge on tests... some educators argue that lobbying for inaccurate diagnoses is particularly common in wealthy communities, where forceful parents search for a psychologist willing to diagnose their child with a learning disorder. (p. 24)

The College Board president acknowledged the risk of falsified diagnoses and stated “we must ensure that extended test-taking time is not granted to students who do not require this accommodation” (p. 24). Given the potential that substantial accommodations could be given to patients and students who do not deserve them, the validity of LD / ADHD evaluations have been under scrutiny in the literature. This scrutiny is well deserved especially since LD and ADHD assessments often rely on self-report and symptom checklists, which are notoriously easy to feign or exaggerate (Harrison, 2006; Sullivan, May, & Galbally, 2007). For example, a study that included non-affected students who were asked to simulate ADHD found they were able to successfully fake symptoms on the ADHD Behavioral Checklist (Quinn, 2002). There was no significant difference between the self-report responses of the ADHD and malingering groups, thus supporting the notion that self-report methods are inadequate as the only information source in ADHD and LD evaluations (Richman et al., 2006; Sullivan, & Richer, 2002).

In a study of ADHD that included cognitive testing but not effort testing, Harrison, Edwards, and Parker (2007) stated “simulators are indistinguishable from those with true ADHD. Students motivated to feign ADHD could easily perform poorly on tests of reading and processing speed, thus allowing them access to academic accommodations” (p. 577). Quinn (2002) found that simulator ADHD malingers reported
using strategies such as: general inattention, ignoring visual and auditory stimuli, commission and omission errors, random responding, hyper-active responding, general fidgety behavior, and slowness of responding. Harrison (2006) noted, “the problem with most of these [effort] tests, however, is that they only measure certain types of exaggeration (e.g., memory)” (p. 3). The author suggests that widely used effort tests, which typically employ memory test formats, may be insensitive to inadequate effort in the context of an ADHD or LD evaluation.

Osmon, Plambeck, Klein, and Mano (2006) outline two hypotheses that help illustrate how effort can influence neuropsychological test scores. First, the domain-specific hypothesis holds that the patient or student performs poorly on tests that are face valid for the types of cognitive deficits attributed to the disorder in question, similar to Lanyon’s (1997) “accuracy of knowledge” conception of malingering. For example, a person who is feigning a memory problem will pick a test that is face valid as a memory test and selectively do poorly on it. Osmon and colleagues (2006) used this hypothesis to develop an effort test designed according to a layperson’s conception of learning disability. Others have also shown support for the “domain-specific” hypothesis (see Heaton, Smith, Lehman, & Vogt, 1978; Nies & Sweet, 1994; Frederick, 1997).

The second hypothesis, general-global, is similar to Lanyon’s (1997) “global signs of lying” conception of poor effort. For example, a person feigning memory problems gives less than his or her best effort on all tests because motivation to perform is less than complete (Green, 2003). This hypothesis also has recently received increased support. For example, Green (2007) showed that failure on an effort test was positively correlated with below average scores on a range of neuropsychological tests among a
forensic litigation sample, including a test that measured finger tapping speed. Sullivan, May, & Galbally (2007) also confirmed this hypothesis in a study of effort among college LD and ADHD evaluations. They found significant correlations between effort test failure and endorsement of psychological symptoms not related to LD or ADHD symptoms.

Osmon and colleagues (2006) developed the Word Reading Test (WRT) based on the domain-specific hypothesis to assess effort in LD and ADHD assessment. The authors compared the WRT with one of the most sensitive and well validated tests of effort, the Word Memory Test (WMT). They also tested the general-global hypothesis against the domain-specific hypothesis. Results from Osmon and colleagues' (2006) study supported the possible effectiveness of WRT error scores, which exceeded the WMT in sensitivity and specificity for LD malingering simulators. The results of the WRT study supported not only the domain-specific hypothesis but also the general-global hypothesis due to the high rate of WMT failure in simulators (discussed in detail below). This study represented a preliminary step in validating the WRT, yet is limited in application until research with a clinical sample, rather than a simulator sample, is conducted.

The present study includes archival data from patients seeking a neuropsychological evaluation. The study used two effort tests similar to those used in the Osmon and colleagues' (2006) study, the WRT and the MSVT (described later). The aim of this study was to test the domain-specific hypothesis against the general-global hypothesis of effort. Failure on the WRT but not the MSVT would support the domain-specific hypothesis. Failure on both effort tests would lend support for general-global
hypothesis. Additionally, failure on an effort test that is face valid for the disorder in question and passing the effort test not face valid for the disorder in question would show support for the domain-specific hypothesis. For example, failing the WRT and passing the MSVT, when the disorder in question is a Learning Disorder, would lend support for the domain-specific hypothesis.

The present study is especially relevant given the potential misallocation of accommodations to individuals with inaccurate diagnoses. The WRT is the only effort test, to date, that is designed specifically for LD evaluations. However, the WRT only has one validation study with a simulator sample. The present study, extending the work of Osmon and colleagues' (2006), used a clinical population to compare the WRT against the MSVT.
REVIEW OF LITERATURE

Effort Testing and Incentives

Patient’s self-reports cannot always be taken at face value especially in the presence of secondary gain. For example, individuals involved in litigation who sustain a mild traumatic brain injury (MTBI) are more likely to show suboptimal effort than those with moderate to severe traumatic brain injury (TBI) (Green, Iverson, & Allen, 1999). Mild head injury is typically defined as a trauma resulting in a period of unconsciousness of 20 minutes or less, a Glasgow Coma Scale score of 13-15, and hospitalization not exceeding 48 hours (Rimel, Giordani, Barth, Boll, & Jane, 1981). Green et al. concluded that patients with definite traumatic brain injuries obtained significantly higher pass rates on effort tests because they had little motivation to exaggerate already verified injuries, when compared to patients with less severe head injuries who had more incentive to give poor effort during testing.

Extensive research on the effects of mild TBI was recently reviewed in a meta-analytic study conducted by the Collaborating Centre Task Force on mild TBI (Carroll et al., 2004). The Task Force selected 128 from over 500 mild TBI articles that were judged methodologically sound. Results indicated the prognosis for adults after mild TBI is complicated by inadequate consideration of the possible confounding effects of other factors including the following: pain, medication, associated injuries, emotional distress, and medicolegal or financial compensation factors. However, the best evidence
consistently suggested there were no cognitive deficits beyond one to three months post injury in the majority of mild TBI cases.

Iverson (2005) concluded similarly in his review of the effects of mild TBI, stating, "in general, excellent recovery from mild TBI is not a well-recognized fact because a small subset of individuals who do not appear to recover well receive a lot of attention in the insurance disability and legal systems, health care system, media (for athletes), and research literature" (p. 303; for similar results, see Binder & Willis, 1991; Flaro, Green, & Robertson, 2007). The evidence clearly suggests the majority of patients with mild TBI do not have lasting neuropsychological deficits beyond three months post injury.

Why then have Binder and Rohling (1996) found that some patients, despite less severe injuries, show more abnormality and disability when overwhelming evidence suggests most recover within three months? The answer can be found in organizational research conducted by Hollenbeck, Ilgen, Phillips, and Hedllund (1994) and Highhouse and Yuce (1996). These researchers found risk taking behavior (i.e., weighing gains more heavily than potential losses) was especially high when subjects had the opportunity of gaining a substantial reward. In our current legal and health care system, impairment and poor prognoses may lead to an increase in financial settlement for the patient, especially for one involved in litigation.

Incentives for poor effort during neuropsychological evaluation have been identified for a variety of circumstances. Richman et al. (2006) noted that patients applying for disability and patients already on disability may alter their behavior on neuropsychological tests in order to obtain and maintain benefits. Patients who are
involved in workers compensation or personal injury litigation also have secondary gain incentive (Green, 2007).

_Malingering_. Neuropsychologists would be impetuous to assume that every person failing an effort test was malingering. The Diagnostic and Statistical Manual of Mental Disorders, 4th edition, Text Revision (DSM-IV-TR; American Psychiatric Association [APA], 2000) defines malingering as the “intentional production of false or greatly exaggerated symptoms for purposes of obtaining some identifiable external reward.” (American Psychiatric Association, 1994, p. 739). There are causes for poor effort during testing other than malingering. For example, it may be the patient’s attempt to alert others to a need for help. It is possible that a dependent person who suffers an actual injury may be led by a stronger authority figure or significant other to maintain his or her symptoms and avoid recovery to reap the maximum benefit possible (Matheson, 1978). In a review of effort testing posted on the National Academy of Neuropsychology website, Williams (1998) wrote:

> Few are the patients who walk through the door with an intention to fake impairment and have carefully planned a strategy to accomplish this goal. Most are fundamentally honest people who are placed under extreme financial pressure to perform worse than their full ability, or they are people with psychological disorders for whom cognitive difficulties are symptoms” (p. 2).

Williams (1998) categorized factitious responding into three general categories. The first category refers to malingering patients (less common) who plan ahead and consciously attempt to appear impaired. The second category refers to patients with somatoform disorders (also less common) whose suboptimal effort usually involves unconscious processes that manifest as physical symptoms without an underlying medical
cause. The third category refers to patients who exaggerate symptoms (more common), and includes patients who have sustained a TBI or other neurological illness who exaggerate their genuine neuropsychological impairment or prolong genuine symptoms that were present soon after the injury but now have recovered (For additional explanations of poor effort, see Braverman 1978; Iverson & Binder, 2000).

Slick, Tan, Strauss and Hultsch (2004) noted that neuropsychologists are cautious in their use of the term malingering. In fact, 41% of the neuropsychologists the authors surveyed reported they rarely use the term malingering and 12% reported they never use it. In a related survey, 81% of neuropsychologists indicated they often or always said that test results suggested or indicated exaggeration if evidence was found. Only 29% reported they often or always stated that test results suggested or indicated malingering, while 24% never stated malingering in a report or professional communication (Sharland & Gfeller, 2007).

Slick, Sherman, & Iverson (1999) outlined diagnostic criteria for Definite, Probable, and Possible Malingering of Neurocognitive Dysfunction, which assist the clinician in defining, communicating, and diagnosing malingering. A clear standard for diagnosing malingering is especially important in neuropsychology where clinical and legal outcomes may depend on accurate diagnoses.

**Base rates.** Base rate studies of suboptimal effort in neuropsychological testing are often organized by specific populations of interest, for example forensic, medical, mood disorder. Larrabee (2005) estimated 40% of patients that were involved in litigation or seeking a disability evaluation failed effort tests. Other estimates of effort
test failure within this population range from 20% to 40%, with the majority of research on base rates within this population coming from extensive work by Green (2003, 2004, & 2007) and others (Binder & Willis, 1991; Gervais, Rohling, Green, & Ford, 2004; Howe, Anderson, Kaufman, Sachs, & Loring, 2007; Mittenberg, Patton, Canyock, & Condit, 2002; Sharland & Gfeller, 2007; Slick Tan, Strauss, & Hultsch, 2004).

Another population consists of patients within a medical context. A survey of neuropsychologists with respect to effort testing indicated neuropsychologists estimated from their clinical work that 5% of patients exaggerated deficits when there was no ongoing litigation or possibility of monetary compensation (Sharland & Gfeller, 2007). In a similar survey study, American Board of Clinical Neuropsychology members estimated symptom exaggeration base rates for medical cases to be approximately 8%, based on a considerable sample size (n = 22,131) (Slick, Tan, Strauss, & Hultsch, 2004). Likewise, Howe, Anderson, Kaufman, Sachs, and Loring (2007) found a 4.8% base rate of suboptimal effort among patients who had established diagnoses of brain injury, leading the authors to conclude that effort testing in a medical context is an integral part of a neuropsychological evaluation.

Individuals undergoing neuropsychological evaluation for academic purposes are another population where symptom validity tests have become increasingly used due to concerns about financial, educational, or prescription incentives mentioned earlier. In her report, aptly named "Adults faking ADHD: You must be kidding!" Harrison (2006) estimated that 20% of adult ADHD referrals significantly exaggerated their symptoms or willfully malingered symptoms of ADHD to receive secondary gain. Sullivan, May, and Galbally (2007) used an effort test in LD and ADHD assessments at a college counseling
center and obtained failure rates at or above those seen in forensic settings. Of the total sample, 22.4% failed the effort test (15.4% of the LD group, and 47.6% of the ADHD group). The ADHD failure rate was above Harrison’s (2006) initial estimate of 20%.

There are currently no other college counseling base rates for comparison. To date, the above articles are the only published studies that include effort test base rates in ADHD and LD evaluations. This research gap suggests a need for continued research; base rates for clinical samples undergoing neuropsychological evaluations for educational purposes would be especially valuable.

**Effort Indicators and Symptom Validity Measures**

Past neuropsychological research has led to development of methods which aid in detection of suboptimal effort, symptom exaggeration, and malingering. Approaches for detection of suboptimal effort in a neuropsychological evaluation can be divided into three categories: patient presentation and history, indices within neuropsychological tests, and independent symptom validity measures. This organization represents a collapsed configuration of Rogers, Harrell, and Liff’s (1993) proposed strategies for detecting feigned neuropsychological deficits.

*Patient presentation and history.* Klonoff and Lamb (1998) reviewed nine case studies in which patient presentation and history explained the presenting dysfunction better than diagnosed neuropathology. They suggested specific characteristics to consider, including medication refusal, coma scale score, refusing a psychological test, length of unconsciousness post injury, emotional distress, atypical response, and salient
pre-injury histories. Other researchers suggested collecting collateral information from sources close to the client in order to enhance detection of suboptimal effort or malingering. Significant others of TBI patients observed significantly more cognitive, emotional-behavioral, and total current problems than did the significant others of malingerers (Sbordone, Seyranian, & Ruff, 2000). However, the suspected malingerers themselves complained of significantly more related problems than even the patients with severe TBI. Researchers also have suggested other approaches to evaluating symptom validity. Iverson and Binder (2000) suggested examining inconsistencies in patient presentation and inconsistencies between medical records and patient self-report, whereas Millis and Volinsky (2001) advised considering pre-existing emotional stress, history of neurological or psychiatric disorder, and chronic social difficulties.

Indices within neuropsychological and psychological tests. Mittenberg, Aguilapuentes, Patton, Canyock, and Heilbronner (2002) proposed profile patterns that would aid in elucidating malingering among widely-used neuropsychological tests such as the Wechsler Adult Intelligence Scale- 3rd edition (WAIS-III; Wechsler, 1997), the Halstead-Reitan Battery (HRB; Reitan & Wolfson, 1985), and the Wechsler Memory Scale – Revised (WMS-R; Wechsler, 1987). Iverson and Binder (2000) suggested looking for discrepancies between obtained and expected scores, as well as discrepancies between known test score relationships (for similar indices see Meyers & Volbrecht, 2003). Although these approaches are abundant in the literature, some have questioned their validity, especially when they are included within a standardized assessment battery that was not originally validated as an effort test (Adams, 2000).
**Development and clinical use of symptom validity measures.** Independent stand-alone instruments used to assess effort in neuropsychological evaluations have become commonly known as Symptom Validity Tests (SVTs). Hartman (2002) proposed six criteria for SVTs that ensure a valid format. First, it must measure a willingness to exert effort and must be insensitive to cognitive dysfunction (sensitivity and specificity). Second, it must appear to the patient to be a realistic measure of cognitive performance. Third, it must have a strong normative basis to satisfy scientific and legal concerns. Fourth, it must be based on validation studies including people without brain injury or suspected of malingering, patient populations, and individuals who are suspected and verified malingerers in actual assessment conditions. Fifth, it should be resistant to coaching. Finally, it must be supported by continuing research. These criteria proposed by the author are stringent, and in his opinion only the Word Memory Test meets these criteria. Other SVT's may only meet some of the above criteria.

It is important to keep in mind that most tests of effort are face valid as memory tests. The reasoning for this is two-fold: First, memory complaints are one of the most common complaints among patients undergoing neuropsychological evaluation (Mittenberg, Agulia-Puentes, Patton, Canyock, & Heilbronner, 2002), and second, is the deceptive nature of memory recognition. In a classic study designed to measure the capacity of memory recall, Shepard (1967) found that subjects were able to correctly identify from memory approximately 600 words, sentences, or pictures when only primed once (for similar results see Nickerson, 1965; Rees, Tombaugh, Gansler & Moczyaski, 1998). Constantinou and McCaffrey (2003) stated, “it has been well documented that
humans possess an astonishing capacity for recognizing previously encountered stimuli” (p. 81). This “astonishing capacity” is not commonly recognized by the general public, which makes memory recognition tests a prime format used to detect suboptimal effort.

Early SVTs were based on the concept of below-chance responding and forced choice responses. The patient is presented with two response options, one correct and one incorrect. Three types of profiles emerge from these forced-choice tests. First is above chance correct responses, in which the individual gives mostly correct answers; second, approximately 50% correct responses; third, below-chance responses, in which the individual gives mostly incorrect answers. If the patient scores at the below-chance level, it can be concluded that they are able to judge correct vs. incorrect responses and show a preference for incorrect answers (for a detailed explanation of forced choice tests, see Frederick, & Speed, 2007). Rogers (1993) wrote “One notable advantage of the SVT over all other strategies is the lack of other viable explanation for below-chance performance” (p. 262). This type of approach has two fundamental benefits. First, below chance responding does not necessitate a need for normative data. Second, the logic of interpreting below chance performance is simple and easily explained to clinicians, judges, and other professionals.

Although forced choice tests are easy to interpret, it has been suggested that they lack sensitivity (Binder, 1993; Frederick, Sarfaty, Johnston, & Powel, 1994; Guilmette, Hart, & Giuliano, 1993). Since so few malingerers perform worse than chance, scores on forced choice tests result in far too many false negative classifications. In response to the poor sensitivity of forced choice tests researchers started to add normative data to these tests. For example, the Test of Memory Malingering (TOMM) measures chance
responding and uses a normative sample to aid in effort assessment (Tombaugh, 1997). Williams (1998) noted that the investigation of symptom validity testing was moving towards normative studies and standard score comparisons of malingering with both brain-injured and control subjects. In accordance with these reviews, recently developed effort tests use normative samples to assess effort (Green, 2003; Green, 2004; Osmon, Plambek, Klein, & Mano, 2006; Schagen, Schmand, Sterke, & Lindeboom, 1997).

It also is important to examine the clinical use of SVT. Recent research has focused on the standards of SVT practice. Slick, Tan, Strauss, and Hultsch (2004) surveyed neuropsychology “experts” and found 79% reported using at least one specialized SVT. The authors suggested using two SVT’s in any case where there is an incentive for suboptimal performance. In a survey by Sharland and colleagues (2007) they found that 56% of respondents often or always include a measure of effort in a neuropsychological evaluation. Many researchers have suggested routinely using a SVT as part of a standard neuropsychological battery (see, Flaro, Green, & Robertson, 2007; Pankratz, 1979; Iverson, 2006; Meyers & Volbrecht, 2003).

Still others have not come to a similar conclusion and leave discretion to the neuropsychologist. For example, Bush et al. (2005) stated, in reference to using SVTs, “when determined by the neuropsychologist to be necessary for the assessment of response validity, administration of specific symptom validity tests are also medically necessary” (p. 425). That same year Bush (2005) published a sample informed consent based on the position paper which included giving a warning to the patient that symptom validity will be assessed:

You are to give your best effort during the testing. This does not mean that you have to get every answer on every problem correct, for no one ever does.
However, you do have to give your best effort. Part of the examination will address the accuracy of your responses, as well as the degree of effort that you give on the tests. (p. 1005)

The authors suggested presenting this information before the evaluation as well as verbally reviewing it with the patient. Other researchers also give warning effort tests will be used in their informed consent. Green, Iverson, & Allen (1999) noted that all patients were warned orally that the two-day neuropsychological assessment included several tests of effort. They were told that if they did not give adequate effort, it would likely be detected.

Despite suggestions to inform patients in advance that effort tests will be used, only 22% of neuropsychology practitioners give patients some type of warning of effort tests and 52% rarely or never give this warning (Sharland & Gfeller, 2007). In support of the current practitioners, Youngjohn, Lees-Haley and Binder (1999) show results suggesting that warning patients about effort test can actually make them more sophisticated malingerers.

Symptom validity tests. I will limit my review of SVTs to those that are currently used by practicing neuropsychologists and researchers. These tests have been identified by extensive surveying of the field (Sharland & Gfeller, 2007; Slick, Tan, Strauss, and Hultsch, 2004).

The first symptom validity test, known as Rey Dot Counting, was developed by Rey (1941) and described by Lezak (1983). The test consists of counting dots on 3 x 5 in. cards while being timed. Although the test is easy to administer and score, it has been criticized for its lack of ongoing validation research (Hartman, 2002).
The Rey 15-Item Test requires memorization of 15 different items arranged in five rows of three characters. Items are shown for 10 seconds then immediate recall is tested (Rey, 1964). It is praised for its ease of administration and scoring but criticized for poor sensitivity and specificity, unclear cutoff scores, and lack of normative data (Williams, 1998; Sbordone, Seyranian, & Ruff, 2000; Hartman, 2000; Rogers, Harrell, & Liff, 1993). Despite these criticisms, it is the second most widely used SVT by neuropsychologists (Slick, Tan Strauss, & Hultsch, 2004).

The Portland Digit Recognition Test (PDRT) introduced by Binder & Willis (1991) has produced a substantial amount of research and normative data compared to the two SVTs described above (Iverson & Binder, 2000). The PDRT uses 5-digit number sequences presented on an index card. The examinee is then instructed to count backward for varying amounts of time. The examinee then chooses incorrect/correct answers on a recognition card. The PDRT has been criticized for its lengthy 72 trials and administration time (Green, Iverson, Allen, 1999). However, it is more face valid as a neuropsychological test than earlier SVTs (Sbordone, Seyranian, & Ruff, 2000; Hartman, 2000).

The Amsterdam Short-Term Memory Test (ASMT), developed and validated by Schagen, Schmand, De Sterke and Lindeboom (1997), is shown to be less affected by client coaching compared to some other SVTs (Gorny, & Merten, 2005). The ASMT was originally published in Dutch and then translated into German and English but has few validation studies. In a review of this test, Lezak, Howieson, and Loring (2004) found three distinct advantages. First, the task uses a two choice recognition format and is less likely to be identified as an effort test by informed patients. Second, the use of high
frequency words for correct answers and low frequency words for incorrect answers increases the likelihood of correct responses from patients giving their best effort.

Frederick (1997) developed the Validity Indicator Profile (VIP), which includes two subcomponents. The first is face valid as a visual discrimination test comparable to the Matrix Reasoning subtest of the WAIS-III. The second subcomponent is a word definition test. The VIP is based on the “domain-specific hypothesis” previously discussed and has been criticized for its poor sensitivity and specificity as well as the overall cost of scoring (Lezak, Howieson & Loring, 2004).

The Victoria Symptom Validity Test (VSVT), introduced by Slick, Hopp, Strauss, and Spellacy (1996), can be administered via computer or index cards. In an initial validation study the VSVT showed promise in detecting suboptimal effort among a litigation sample. However, little research has examined its psychometric properties (Lezak, Howieson & Loring, 2004) and it is only used “often” by 7% of neuropsychologists (Sharland & Gfeller, 2007).

The Test of Memory Malingering (TOMM) developed by Tombaugh (1997) is one of the most widely used SVTs (Sharland & Gfeller, 2007; Slick, Tan Strauss, & Hultsch, 2004). The test consists of 50 pictures presented twice and tested three times. It is face valid as a three trial visual memory test. Some researchers have criticized the TOMM for its low sensitivity to suboptimal effort among actual patients (Gervais, Rohling, Green & Ford, 2004), yet it has been found useful among a variety of clinical populations (Lindem et al., 2003).

The Word Memory Test (WMT) created by Green, Allen, and Astner (1996) is perhaps one of the most often used SVTs to date. The test consists of 20 word pairs with
strong semantic associations and is face valid as a memory test. Comparison groups and base rates have been gathered on children with and without neurological disease; college students; custody seeking parents; non-English speaking adults; adult control; adults seeking disability evaluations; adults simulating impairment; adults with depression, mild head injury, and TBI; as well as extensive systemic and neurological groups (Green, 2003; Sullivan, May, & Galbally, 2007).

Advantages of the WMT include oral or computer administration, effort and actual memory scores, and a gradient of difficulty across measures (Lezak, Howieson & Loring, 2004; Green, Iverson & Allen, 1999). The WMT has been found to be more sensitive to suboptimal effort than the TOMM (Gervais, Rohling, Green, & Ford, 2004). In his review and critique of eight currently available SVTs, Hartman (2002) suggested that only the WMT satisfied all of his criteria as a stand-alone test of effort. According to Hartman, it is the only test that is robust to coaching, although this assumption was tested by Gorny and Merten (2005) who found that the Amsterdam Short-term Memory Test was more robust to simulator coaching than the WMT.

Critiques of the WMT have been rare. However, Hartman (2002) noted that the WMT manual was poorly organized and confusing. Bowden, Shores, and Mathias (2006) used the WMT with a sample of children and adults in medical settings and concluded that the test is an invalid measure of effort. In rebuttal Flaro, Green and Robertson (2007) reviewed Bowden and colleagues data and pointed out that the study did not follow standard administration, had a small sample, did not differentiate adult scores versus children, and based their results on the administration of only one of seven subtests.

Green (2004) released a second effort test that is similar to the WMT. This new
test, called the Medical Symptom Validity Test (MSVT), is a simplified version of the WMT (Richman et al., 2006). The MSVT has half as many words to remember as the WMT and includes only five of the seven WMT subtests. Ten word pairs representing common objects are presented over two trials. Following the presentations, Immediate Recognition memory (IR) is tested. After a 10-minute delay, Delayed Recognition memory (DR) is tested, followed by a Paired-Associates trial (PA) where the first word of each pair is presented and the ability to recall the second word is assessed. Finally, there is a Free Recall trial (FR). In addition to memory performance, a consistency variable (CNS) is calculated to reflect recall consistency across tasks (Howe, Anderson, Kaufman, Sachs, & Loring, 2007).

The MSVT is intended to be used by other professionals who measure cognitive functioning such as physicians and health professionals who evaluate disability. In demonstrating the MSVT’s relative insensitivity to cognitive abilities, Chafetz, Abrahams, & Kohlmaier (2007) demonstrated that even children with Full Scale IQ scores less than 70 can pass the MSVT. Merten, Green, Henery, Blaskewitz, and Brockhaus (2005) investigated a German language version of the MSVT in an experimental study using simulators and a control group. The authors reported 100% sensitivity and specificity between the two groups. However, as noted earlier, care must be taken in interpreting studies using simulators. Others have used the MSVT with a clinical sample showing results and base rates similar to previous validation studies of the WMT (Howe, Anderson, Kaufman, Sachs & Loring, 2007).

The final SVT I will discuss takes a different approach than those previously described. In concordance with the “domain-specific” hypothesis of effort, Osmon,
Plambeck, Klein and Mano (2006) introduced an SVT designed to measure symptom exaggeration of feigned learning disorders. Osmon and colleagues (2006) give a brief description of the test:

A word is presented on a computer screen for a brief duration, then two words are immediately presented on a subsequent screen without delay and without backward mask, such that the task would not likely tax word reading skills even in poor readers. The choice is between the actual target word and a foil that contains a similar but incorrect choice. Thus, foils consist of choices with characteristics that might likely be attributed to language problems, such as mirror letters (develop vs. bevelop), homophones (too vs. two), and additions/deletions of letters (through vs. thorough). Additionally, task instructions indicated that speedy but accurate performance is important and both reaction time and error scores are included. (p. 316)

The WRT is the first symptom validity test designed specifically for use within a LD evaluation. It relies on the “domain-specific” hypothesis, which holds “that individuals’ difficulty on effort tests arises from suboptimal effort on specific tests that are face valid for the types of cognitive deficits attributed to the disorder in question by laypersons” (Osmon, Plambeck, Klein and Mano, 2006, p. 316). Thus, persons feigning a learning disability will give suboptimal effort “specifically” on tests that are face valid for the associated disorder (e.g., reading and math tests). The notion of face validity for the disorder in question is one of the preferred criteria for SVTs set forth by Hartman (2002).

The WRT introduced a relatively new aspect of effort testing in that it measured response latencies. This is the only test reviewed in which response time is used as an effort measure. Multiple researchers over the years have suggested response time to be a viable effort variable (Rose, Hall, Szalda-Petree & Bach, 1998; Goebe, 1983; Iverson, 1995; Beetar & Williams, 1995).

The initial validation research of the WRT with LD simulators and a control group showed a 90.32% sensitivity and a 100% specificity for the WRT mean correct
responses, which outperformed the WMT (96% specificity and 65.52% sensitivity).
However, reaction time did not significantly discriminate between reading and speed simulators as hypothesized by the authors and did not outperform the WMT. The authors conclude, “the WRT appears to be a potentially clinically useful measure to detect effort in adult learning disability populations. Further research in an actual clinical population is warranted before using the instrument in a clinical setting” (p. 322).

Hypotheses of the Study

The present study sought to extend Osmon and colleagues' (2006) research by collecting base rate data in a clinical group, as well as test the general-global and domain-specific hypotheses. The MSVT is a well-validated effort test, which is face valid as a memory test. The WRT is a newly developed effort test designed to assess effort among a learning disability population.

Failure on both effort tests would lend support for the general-global hypothesis. In contrast, failure on an effort test that is face valid for the disorder in question and passing the effort test that is not face valid for the disorder in question would demonstrate support for the domain-specific hypothesis. For example, failing the WRT and passing the MSVT, when the disorder in question is a Learning Disorder, would lend support for the domain-specific hypothesis.
METHOD

Subjects

Subjects were 30 participants referred to a university doctoral clinical psychology training and research clinic for neuropsychological evaluation; the clinic services the Portland, Oregon metropolitan area. They ranged in age from 18 to 55 years. Of these cases, 17 (57%) were referred for assessment of ADHD or LD and 13 (43%) for assessment of general cognitive problems and Asperger’s syndrome evaluation. The sample consisted of 21 (70%) college or university referrals and 9 (30%) medical or psychological referrals. Fifty-three percent of the sample was female and 47% were male. Eighty-seven percent of the sample was right hand dominant and 89% were Caucasian. Ten percent of the sample had a history of head injury that included loss of consciousness. All participants demonstrated cognitive ability above what is typically required to successfully pass both the MSVT and the WRT as demonstrated by WAIS-III characteristics, education, and age of the sample, see Table 1. (Green, 2004; Osmon, Plambeck, Klein & Mano, 2006).
Table 1

*Age, Education and WAIS-III Indices of the Sample*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>31.60</td>
<td>10.32</td>
</tr>
<tr>
<td>Education Years</td>
<td>14.17</td>
<td>2.12</td>
</tr>
<tr>
<td>WAIS-III Verbal IQ</td>
<td>107.07</td>
<td>16.57</td>
</tr>
<tr>
<td>WAIS-III Performance IQ</td>
<td>104.97</td>
<td>15.10</td>
</tr>
<tr>
<td>WAIS-III Full Scale IQ</td>
<td>106.83</td>
<td>16.07</td>
</tr>
<tr>
<td>WAIS-III VCI</td>
<td>112.10</td>
<td>14.77</td>
</tr>
<tr>
<td>WAIS-III POI</td>
<td>108.07</td>
<td>17.467</td>
</tr>
<tr>
<td>WAIS-III WMI</td>
<td>96.33</td>
<td>18.01</td>
</tr>
<tr>
<td>WAIS-III PSI</td>
<td>97.80</td>
<td>12.36</td>
</tr>
</tbody>
</table>

*Note.* WAIS-III = Wechsler Adult Intelligence Scale 3rd Edition; VCI = Verbal Comprehension Index; POI = Perceptual Organization Index; WMI = Working Memory Index; PSI = Processing Speed Index; SD = standard deviation.

*Procedure*

Participants consented to have their clinical data included in a research database and were evaluated between September 2007 and June 2008. IRB approval was obtained from the institution (Pacific University IRB Approval #046-08). Data was entered into a de-identified computer database.
The original windows version of the MSVT and the WRT effort tests were used in standard administration as described by Green (2004) and Osmon and colleagues (2006). Suggested cutoff scores for the MSVT are 85% or below on Immediate Recall (IR), Delayed Recall (DR), and Consistency (CNS). Suggested WRT Total Errors cutoff score is four. These cutoffs were the criteria for failure on each test. The Wechsler Adult Intelligence Scale, 3rd Edition (WAIS-III) was administered as part of a comprehensive neuropsychological test battery.

Neuropsychological test scores, effort test scores (MSVT and WRT), and non-identifying demographic data were collected from 30 archival neuropsychological assessments. These assessments took place over a 1-year period. The university clinic where the evaluations were done receives cognitive evaluation referrals from surrounding colleges and universities as well as from other psychological and medical practitioners. Participants were assessed during four testing sessions that occurred on separate days. Clinicians conducting the assessments were graduate or doctoral internship students. All students were supervised by a licensed neuropsychologist.

Patient files were evaluated to determine if patients had potential secondary gain. Potential secondary gain was considered present when the patient could have received one of the following if they met criteria for a diagnosis: accommodations in school, vocational rehabilitation assistance, disability assistance, or social security benefits.

The Statistical Program for Social Sciences, version 15.0 (SPSS, 2006), was used to analyze the data.
RESULTS

Of the 30 cases analyzed 1 (3%) failed the WRT only, 3 (10%) failed the MSVT only, and 2 (7%) failed both effort tests. Six (20%) of the participants were found to have given suboptimal effort as measured by at least one effort test failure.

MSVT results showed 5 of the 30 cases failed according to the cutoff scores presented in the manual (Green, 2004). Of the 5 participants, 3 participants scored below the cutoff for Immediate Recall and Delayed Recall. All 5 participants scored below the Consistency cutoff and were Caucasian. Group means and standard deviations from the MSVT are presented in Table 2. A chi-square analysis was conducted to evaluate the relationship between referral source (university or other), demographic variables, diagnostic group (ADHD, LD, Anxiety Disorder) and effort test failure on the MSVT. The results of the test were not significant for gender, diagnostic group and referral source, see Table 3). An analysis of variance (ANOVA) was conducted using the Bonferroni approach to control for Type I error for multiple comparisons. The adjusted probability level for significance was .008 (.05/6). There was not a significant difference between WAIS-III indices, age, education, and effort test failure on the MSVT (see Table 4).
Table 2

*Group Means and Standard Deviations for Percent Correct on the MSVT Subtests*

<table>
<thead>
<tr>
<th>WMT Score</th>
<th>Failed MSVT n=5</th>
<th>Passed MSVT n=25</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR</td>
<td>88.00 6.71</td>
<td>98.41 3.58</td>
</tr>
<tr>
<td>DR</td>
<td>87.00 10.37</td>
<td>98.64 3.51</td>
</tr>
<tr>
<td>CNS</td>
<td>79.00 10.84</td>
<td>100.00 .00</td>
</tr>
<tr>
<td>PA</td>
<td>67.00 22.24</td>
<td>97.73 4.29</td>
</tr>
<tr>
<td>FR</td>
<td>80.00 15.81</td>
<td>77.50 11.83</td>
</tr>
</tbody>
</table>

*Note.* MSVT = Medical Symptom Validity Test; IR = Immediate Recall; DR = Delayed Recall; CNS = Consistency; PA = Paired Associates; FR = Free Recall; SD = standard deviation
Table 3

Chi-Square Results for those failing the Medical Symptoms Validity Test

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pass MSVT (n = 25)</th>
<th>Fail MSVT (n = 5)</th>
<th>DF</th>
<th>( \chi^2 )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td>.20</td>
<td>.65</td>
</tr>
<tr>
<td>Male</td>
<td>11 (44%)</td>
<td>3 (60%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>14 (56%)</td>
<td>2 (40%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic Group</td>
<td></td>
<td></td>
<td></td>
<td>.40</td>
<td>.82</td>
</tr>
<tr>
<td>ADHD</td>
<td>5 (20%)</td>
<td>2 (40%)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD</td>
<td>4 (16%)</td>
<td>1 (10%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety Disorder</td>
<td>1 (4%)</td>
<td>2 (40%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>15 (60%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Referral Source</td>
<td></td>
<td></td>
<td></td>
<td>.18</td>
<td>.18</td>
</tr>
<tr>
<td>University</td>
<td>17 (68%)</td>
<td>4 (80%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>8 (32%)</td>
<td>1 (20%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possibility of</td>
<td></td>
<td></td>
<td></td>
<td>.20</td>
<td>.85</td>
</tr>
<tr>
<td>Secondary Gain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>14 (56%)</td>
<td>2 (40%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>11 (44%)</td>
<td>3 (60%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* \* = Significant (.05)
Table 4

Analysis of Variance for Effort Test Failure on the Medical Symptom Validity Test

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pass MSVT (n=25)</th>
<th>Fail MSVT (n=5)</th>
<th>DF</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>SD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>32.40</td>
<td>27.60</td>
<td>19</td>
<td>1.35</td>
<td>.32</td>
</tr>
<tr>
<td>Education</td>
<td>14.52</td>
<td>12.4</td>
<td>7</td>
<td>2.43</td>
<td>.05</td>
</tr>
<tr>
<td>WAIS-III VCI</td>
<td>112.92</td>
<td>108</td>
<td>20</td>
<td>2.36</td>
<td>.09</td>
</tr>
<tr>
<td>WAIS-III POI</td>
<td>108.76</td>
<td>104.60</td>
<td>18</td>
<td>.71</td>
<td>.74</td>
</tr>
<tr>
<td>WAIS-III WMI</td>
<td>96.16</td>
<td>97.20</td>
<td>20</td>
<td>.57</td>
<td>.86</td>
</tr>
<tr>
<td>WAIS-III PSI</td>
<td>98.52</td>
<td>94.20</td>
<td>14</td>
<td>.50</td>
<td>.89</td>
</tr>
</tbody>
</table>

*Note. WAIS-III = Wechsler Adult Intelligence Scale Third Edition; VCI = Verbal Comprehension Index; POI = Perceptual Organization Index; WMI = Working Memory Index; PSI = Processing Speed Index; WAIS-III indices group by intervals of 10 points. * = Significant (.008)

WRT results showed 3 of the 30 cases failed according to the cutoff scores recommended by Osmon & Colleagues (2006). All 3 participants were Caucasian, referred from a university and had the possibility of secondary gain. A chi-square analysis was conducted to evaluate the relationship between gender and diagnostic group (ADHD, LD) and effort test failure on the WRT. The results of the test were not significant, see Table 5). An analysis of variance (ANOVA) was conducted using the Bonferroni approach to control for Type I error for multiple comparisons. The adjusted probability level for significance was .008 (.05/6). There was not a significant difference
between WAIS-III indices, age, education, and effort test failure on the WRT (see Table 6).
Table 5

*Chi-Square Results for Effort Test Failure on the Word Reading Test*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pass WRT (n = 27)</th>
<th>Fail WRT (n = 3)</th>
<th>DF</th>
<th>$\chi^2$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>14 (52%)</td>
<td>1 (33%)</td>
<td>1</td>
<td>.33</td>
<td>.56</td>
</tr>
<tr>
<td>Female</td>
<td>13 (48%)</td>
<td>2 (67%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADHD</td>
<td>6 (22%)</td>
<td>2 (67%)</td>
<td>1</td>
<td>.50</td>
<td>.78</td>
</tr>
<tr>
<td>LD</td>
<td>4 (15%)</td>
<td>1 (33%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>17 (63%)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Referral Source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>18 (67%)</td>
<td>3 (100%)</td>
<td>1</td>
<td>2.6</td>
<td>.10</td>
</tr>
<tr>
<td>Other</td>
<td>9 (33%)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possibility of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary Gain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13 (48%)</td>
<td>3 (100%)</td>
<td>1</td>
<td>.13</td>
<td>.71</td>
</tr>
<tr>
<td>No</td>
<td>14 (52%)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* * = Significant (.05)
Table 6

Analyses of Variance for Effort Test Failure on the Word Reading Test

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pass WRT</th>
<th>Fail WRT</th>
<th>DF</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n=27)</td>
<td>(n=3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>32.40</td>
<td>10.48</td>
<td>27.60</td>
<td>10.07</td>
<td>19</td>
</tr>
<tr>
<td>Education</td>
<td>14.52</td>
<td>2.04</td>
<td>12.40</td>
<td>1.67</td>
<td>7</td>
</tr>
<tr>
<td>WAIS-III VCI</td>
<td>112.92</td>
<td>14.67</td>
<td>108.00</td>
<td>16.33</td>
<td>20</td>
</tr>
<tr>
<td>WAIS-III POI</td>
<td>108.76</td>
<td>18.57</td>
<td>104.60</td>
<td>11.17</td>
<td>18</td>
</tr>
<tr>
<td>WAIS-III WMI</td>
<td>96.16</td>
<td>17.413</td>
<td>97.20</td>
<td>23.36</td>
<td>20</td>
</tr>
<tr>
<td>WAIS-III PSI</td>
<td>98.52</td>
<td>12.49</td>
<td>94.20</td>
<td>12.36</td>
<td>14</td>
</tr>
</tbody>
</table>

Note. WAIS-III = Wechsler Adult Intelligence Scale Third Edition; VCI = Verbal Comprehension Index; POI = Perceptual Organization Index; WMI = Working Memory Index; PSI = Processing Speed Index; WAIS-III indices group by intervals of 10 points.

* = Significant (.008)

Using the cutoff score of four or more WRT errors suggested by Osmon and colleagues (2006) indicated that two of the three WRT Error failures concurred with the MSVT failure classification. However, using a cutoff score of three or more WRT Errors showed that three of four error failures were classified more similarly to the MSVT failure classification (see Table 7). WRT Reaction Time was compared for those who passed and those who failed the WRT, see table 8 for results.
### Table 7

**Effort Failure Overlap with Cutoff Scores for the WRT Mean Errors**

<table>
<thead>
<tr>
<th>WRT Errors Cutoff</th>
<th>Percent of Effort Failure Concordance with the MSVT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>75.0%</td>
</tr>
<tr>
<td>4</td>
<td>66.6%</td>
</tr>
</tbody>
</table>

#### WRT Error: 4 or more

<table>
<thead>
<tr>
<th>Pass WRT</th>
<th>Fail WRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSVT</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>3</td>
</tr>
</tbody>
</table>

#### WRT Error: 3 or more

<table>
<thead>
<tr>
<th>Pass WRT</th>
<th>Fail WRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSVT</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>2</td>
</tr>
</tbody>
</table>

**Note.** WRT = Word Reading Test; MSVT = Medical Symptom Validity Test
Table 8

ANOVA Results for Errors and Reaction Time on the Word Reading Test

<table>
<thead>
<tr>
<th>Group</th>
<th>Passed WRT</th>
<th>SD</th>
<th>Failed WRT</th>
<th>SD</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Errors</td>
<td>.46</td>
<td>.65</td>
<td>4.50</td>
<td>1.73</td>
<td>.001**</td>
</tr>
<tr>
<td>Mean RT</td>
<td>.97</td>
<td>.21</td>
<td>1.61</td>
<td>.51</td>
<td>.025*</td>
</tr>
</tbody>
</table>

Note. WRT = Word Reading Test, RT = reaction time in seconds, SD = standard deviation. Failure classification was based on a cutoff score of 3 or above. *p < .05, **p < .01

Six of the 30 cases failed at least one effort test according to the cutoff scores presented by the authors (Green, 2004; Osmon & Colleagues, 2006). All 6 participants were Caucasian. A chi-square analysis was conducted to evaluate the relationship between gender, diagnostic group (ADHD, LD, and Anxiety), possibility of secondary gain, referral source and effort test failure at least one effort test. The results of the test were not significant, see Table 9). An analysis of variance (ANOVA) was conducted using the Bonferroni approach to control for Type I error for multiple comparisons. The adjusted probability level for significance was .008 (.05/6). There was not a significant difference between WAIS-III indices, age, education, and effort test failure on at least one effort test (see Table 10).
Table 9

*Chi-Square Results for Effort Test Failure on Either the MSVT or the WRT*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pass both effort tests (n = 24)</th>
<th>Fail at least one effort test (n = 6)</th>
<th>DF</th>
<th>$\chi^2$</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td>1</td>
<td>.00</td>
<td>.99</td>
</tr>
<tr>
<td>Male</td>
<td>11 (46%)</td>
<td>3 (50%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>13 (54%)</td>
<td>3 (50%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic Group</td>
<td></td>
<td></td>
<td>2</td>
<td>.67</td>
<td>.88</td>
</tr>
<tr>
<td>ADHD</td>
<td>5 (21%)</td>
<td>3 (50%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD</td>
<td>7 (29%)</td>
<td>1 (17%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>12 (50%)</td>
<td>2 (33%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Referral Source</td>
<td></td>
<td></td>
<td>1</td>
<td>2.66</td>
<td>.10</td>
</tr>
<tr>
<td>University</td>
<td>16 (67%)</td>
<td>5 (83%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>8 (33%)</td>
<td>1 (17%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possibility of</td>
<td></td>
<td></td>
<td>1</td>
<td>.16</td>
<td>.68</td>
</tr>
<tr>
<td>Secondary Gain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13 (54%)</td>
<td>3 (50%)</td>
<td>1</td>
<td>.00</td>
<td>.99</td>
</tr>
<tr>
<td>No</td>
<td>11 (46%)</td>
<td>3 (50%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* * = Significant (.05)
Table 10

**Analysis of Variance for Effort Test Failure on Either the MSVT or the WRT**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pass WRT and MSVT (n = 27)</th>
<th>Fail WRT or MSVT (n = 6)</th>
<th>DF</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (n = 27)</td>
<td>Mean (n = 6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>32.67 10.62</td>
<td>27.33 9.03</td>
<td>19</td>
<td>1.63</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td>14.54 2.09</td>
<td>12.67 1.63</td>
<td>7</td>
<td>.23</td>
</tr>
<tr>
<td></td>
<td>WAIS-III VCI</td>
<td>113.25 14.89</td>
<td>107.50 14.65</td>
<td>20</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>WAIS-III POI</td>
<td>108.75 18.97</td>
<td>105.33 10.15</td>
<td>18</td>
<td>.92</td>
</tr>
<tr>
<td></td>
<td>WAIS-III WMI</td>
<td>96.25 17.78</td>
<td>96.67 20.94</td>
<td>20</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>WAIS-III PSI</td>
<td>98.33 12.72</td>
<td>95.67 11.62</td>
<td>14</td>
<td>.47</td>
</tr>
</tbody>
</table>

Note. WAIS-III = Wechsler Adult Intelligence Scale Third Edition; VCI = Verbal Comprehension Index; POI = Perceptual Organization Index; WMI = Working Memory Index; PSI = Processing Speed Index; WAIS-III indices group by intervals of 10 points. * = Significant (.008)

Correlation coefficients were computed for the five subtests of the MSVT and the Reaction Time and Errors of the WRT. Using the Bonferroni approach to control for Type I error across the 10 correlations, a $p$ value of less than .005 (.05/10) was required for significance. The results of the correlation analysis presented in Table 11 show that 6 of the 10 correlations were statistically significant and greater than or equal to .50. The WRT Reaction Time was not significantly correlated with MSVT: DR, CNS, PA, FR or WRT Errors, yet was significant with MSVT IR. The highest correlation between the two effort tests was between the WRT Errors and the MSVT CNS at $r = -.64$. 
Table 11

*Correlations between MSVT and WRT Scores*

<table>
<thead>
<tr>
<th></th>
<th>WRT RT</th>
<th>WRT Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSVT IR</td>
<td>-.50*</td>
<td>-.51*</td>
</tr>
<tr>
<td>MSVT DR</td>
<td>-.20</td>
<td>-.63*</td>
</tr>
<tr>
<td>MSVT CNS</td>
<td>-.27</td>
<td>-.64*</td>
</tr>
<tr>
<td>MSVT PA</td>
<td>-.02</td>
<td>-.50*</td>
</tr>
<tr>
<td>MSVT FR</td>
<td>-.14</td>
<td>-.35</td>
</tr>
</tbody>
</table>

*Note. WRT = Word Reading Test; RT = Reaction Time; MSVT = Medical Symptom Validity Test; IR = Immediate Recall; DR = Delayed Recall; PA = Paired Associates; FR = Free Recall*

*p < .05
DISCUSSION

The purpose of this study was to test the domain-specific hypothesis against the general-global hypothesis of effort. At first glance, the results of this study support both the general-global and the domain-specific hypotheses of effort. Of the six cases that failed one or more of the effort tests, Two (33%; using a cutoff of 4 or more WRT Errors) failed both tests, which suggests overlap exists between effort measurements but not complete overlap. However, using a WRT Errors cutoff score of 3 shows a higher concordance between the WRT and MSVT (50%). Further research needs to be done to best determine sensitivity and specificity of the 3 or more WRT Error cutoff score. In addition, 6 of the 10 subtest correlations were statistically significant and greater than or equal to .50.

Failure rate for the MSVT and the WRT was not significantly different for patients referred from different sources (e.g., university or independent practice psychologist). This was somewhat unexpected for the WRT as it was designed specifically to detect suboptimal effort among patients referred for LD evaluation. Since referrals for LD evaluation came from university counseling centers we might have expected that failure on the WRT would be related to the referral source.

These results are inconsistent with Green’s (2007) research showing SVT failure is associated with lower test scores across multiple neurocognitive domains. There were no differences between the WAIS-III scores of those who passed and failed at least one effort test.
Results are consistent with the use of effort tests face valid for memory to be used within LD/ADHD evaluations (Sullivan, May, & Galbally, 2007). Four of the 5 subjects who failed the MSVT were referred from a university for a LD assessment; thus SVTs that are face valid as memory tests may be helpful in detecting poor effort during a LD evaluation.

The present study supports the validity of the WRT as an effort test. Error scores were significantly correlated with MSVT IR, DR, and CNS scores. However, WRT Reaction Time was only correlated to MSVT IR. This result is similar to Osmon and colleagues' (2006) study on student simulators. Remarkably, Reaction Time means and standard deviations from Osmon and colleges’ (2006) simulator sample were nearly identical to the present study. This lends support to the hypothesis that reaction time may also be useful as an indicator of effort.

The most significant limitation of this study was the small sample size. Only 6 of the 27 subjects failed one or both of the effort tests making some parametric and nonparametric analyses less powerful. The current study was also limited in its applicability to other learning disorders such as math and nonverbal learning disorders. The WRT and MSVT are distinctively language-based and are not face valid for learning disorders in other cognitive domains. Unaccounted for order effects of the effort tests and neuropsychological tests could also be considered a limitation. As noted by Osmon and colleagues (2006), order effects may influence outcome, although Guilmette, Whelihan, Hart, Sparadeo, and Buongiorno (1996) found nonsignificant results in a similar study for the order of effort tests versus neuropsychological tests.
Malingering as well as multiple psychological factors is a possible explanation for effort test failure (discussed in the literature review). The sample in the present study was a clinical population, making it impossible to know with confidence why particular participants failed effort tests. No participants admitted to giving suboptimal effort. The present study did not show significant differences on effort tests between those who had a secondary gain incentive and those who did not; thus suggesting that effort tests should be given to all patients and not just those with potential secondary gain present. In addition, no participants showed symptoms of psychological disorders that may mimic suboptimal effort results. Thus the results of this study do not shed much light on why subjects failed SVT, but they do indicate some factors that unlikely were the cause. Continued research in this area will help clarify the meaning of effort test failure.

As the current study was archival, some variables that may have been useful were not initially reported. Possible coding for medication seeking may have shown significant differences among effort test failure; however, this information was inconsistently given in the participants' history.

Recommendations include further research on the psychometric properties of both the WRT and the MSVT. The WRT and MSVT should be assessed for utility with other aspects of learning disorder such as non-verbal LD and mathematics LD. Given the support for the general-global hypothesis, research could focus on the use of SVTs among other populations where memory may not be the presenting symptom yet secondary gain is still present.
REFERENCES


